**Problem Challenge:**

The Company named Tosco and Spency has two kinds of customers, one who buys few expensive products and the other who buys usually many cheap products.

The company wants to differentiate its offer and marketing strategy based on these two kinds of customers.

The company wants a machine learning model that can predict future customers' class, whether he/she belongs to the first class or second.

**Proposed Solution:**

Machine learning has so much to do with today's businesses. Companies use it to predict useful information based on their recorded historical data.

By using machine learning procedures, we can have a better understanding of what we are trying to achieve and how we should make better decisions to achieve success in the market.

**Predictive Task:**

As our problem is to identify between two classes of our customers, so this type of problem falls into classification problems. We will be doing classification tasks.

**Possibly Informative Features:**

It would have been better if we have been provided with the transaction data of our customers, at which time they made their transaction, and how much money each transaction contained. These features would have been useful to learn a pattern between our customers and eventually our predictions would have improved.

**Learning Procedures:**

**Decision Tree Classifier:**

We will be testing this model if the predictions are good, then we are going to use this model to make predictions on our test dataset. The reason why we choose this model is the each of use. We do not have to do any kind of normalization of our dataset if we are using the Decision Tree Classifier model.

**Random Forest Classifier:**

We will also be testing this model if it performs better than the other two models then, we will choose this model. The main reason we choose this model is that It provides higher accuracy through cross-validation. A random forest classifier will handle the missing values and maintain the accuracy of a large proportion of data. If there are more trees, it won't allow over-fitting trees in the model.

**Logistic Regression Classifier:**

This is the most widely used classifier model we will also be testing this model to find out if it performed better than the other two models then we will choose based on the accuracy score of the model. The reason we choose this model is its simplicity and popularity for binary classification problems.

**Evaluating Performance:**

We will choose an accuracy matric to evaluate each model and then we will compare them on accuracy score. Accuracy matric is the most commonly used evaluation metric for classification models.

**Conclusion:**

Hence, we can conclude that it is possible to apply machine learning procedures to predict the customers' classes from our dataset. But the accuracy of our predictions entirely depends on what algorithm we choose and how we preprocess our data before feeding it to our models. We will do our best to apply different machine learning techniques, and compare them together and select the most accurate one.